

# Comparison of daily rainfall forecasting using multilayer perceptron neural network model

Mazwin Arleena Masngut<sup>1</sup>, Shuhaida Ismail<sup>2</sup>, Aida Mustapha<sup>3</sup> Suhaila Mohd Yasin<sup>4</sup>

<sup>1,2</sup>Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh, Muar, Johor, Malaysia

<sup>3,4</sup>Faculty of Computer Science and Information Management, University Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, Johor, Malaysia

## Article Info

### Article history:

Received Feb 18, 2020

Revised Apr 18, 2020

Accepted May 2, 2020

### Keywords:

Artificial neural network

Autoregressive

Daily rainfall

Forecasting performance measurement

## ABSTRACT

Rainfall is important in predicting weather forecast particularly to the agriculture sector and also in environment which gives great contribution towards the economy of the nation. Thus, it is important for the hydrologists to forecast daily rainfall in order to help the other people in the agriculture sector to proceed with their harvesting schedules accordingly and to make sure the results of their crops would be satisfying. This study is set to forecast the daily rainfall future value using ARIMA model and Artificial Neural Network (ANN) model. Both method is evaluated by using Mean Absolute Error (MAE), Mean Forecast Error (MFE), Root Mean Squared Error (RMSE) and coefficient of determination ( $R^2$ ). The results showed that ANN model has outperformed results than ARIMA model. The results also showed ANN has under-forecast the daily rainfall data by 2.21% compare to ARIMA with over-forecast of -3.34%. From this study, it shows that the ANN (6,4,1) model produces better results of MAE (8.4208), MFE (2.2188), RMSE (34.6740) and  $R^2$  (0.9432) compared to ARIMA model. This has proved that ANN model has outperformed ARIMA model in predicting daily rainfall values.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Shuhaida Ismail,  
Faculty of Applied Scienced and Technology,  
University Tun Hussein Onn Malaysia.  
Pagoh, Muar, Johor, Malaysia.  
Email: shuhaida@uthm.edu.my

## 1. INTRODUCTION

Back in 2017, half of Pulau Pinang had submerged with floodwaters particularly in island as well as in mainland. It had caused a lot of damages and fatal in some areas. It is needed to find a suitable method that can predict long term prediction of rainfall. From finding a suitable method will be able to assist the authorities or certain parties to be well-prepared and make plans to prevent these water-related problems from happening.

Other than that, rainfall forecasting is very important in agriculture fields which can also help in decision making and performing strategic planning. The ability to predict and forecast rainfall quantitatively can help crop planting decisions, reservoir water resource allocation, traffic control, the operation of sewer systems and confronting water-related problems such as flood and drought [1].

Previous researchers have shown an increased interest in model development of time series in using rainfall data. There were several attempts in forecasting rainfall data using various techniques and methods in which can produce a well development model. Forecasting method has become very popular

among researchers due to the development of the data collection tools, computation methods and computer software to perform the analysis. There are several methods can be used to perform the time series forecasting analysis which are Multiple Linear Regression, Genetic Algorithm, Support Vector Machine (SVM), and Fuzzy Logic [2-3].

Many researchers used ANN as the basic idea of model combination in forecasting rainfall models to capture different patterns of the data. The parallel distributed processing architecture of ANN has proved to be a very powerful computational tool which is now being used in several fields to model the dynamic processes successfully including the rainfall [4-5]. However, a major problem with this kind of application is that many researchers have tried to compare the performance of the ARIMA and ANN models, but their results were different from each other.

An extensive comparative study between ARIMA and ANN model were carried out by many past researchers in order to investigate the most suitable models that can be used to forecast rainfall data. ANN is a non-linear model which have been widely used for resolving forecasting problem as identified by [6-11]. Previous researchers such as in [12-13] had performed a comparative analysis between ARIMA and ANN model to determine the appropriate models for rainfall forecasting. They conclude that ANN method was appropriate in forecasting rainfall which outperforms ARIMA methods. Similarly, [14] also found that ANN can outperforms the ARIMA models. They confirmed that ANN methods are capable in modelling the complex Rainfall-runoff relationship.

Artificial Neural Networks were widely used in many different fields such as digital image [15], fault detection [16], gold price forecasting [17], healthcare [18] and many more. For an instance, [19] used ANN to forecast daily rainfall on Turkey to observe wavelet of the ANN model. As a result, a close estimation for daily rainfall peaks was observed. Similarly, a study conducted by [20] presented a successful integration of wavelet and ANN for monthly rainfall predictions in India. Furthermore, [21] used Back-Propagation (BP) of ANN to model an hourly rainfall runoff. The results show that BP of ANN can perform satisfactorily and proved to be superior model which presented an acceptable ability to find the relationship between rainfall and runoff using only rainfall data and runoff data.

Studies by [22-23] examined the forecasting models from different techniques to make a comparison in identifying the best model for prediction of rainfall. The author found out that ANN approach is better than any models as it can analyses a non-linear behaviour pattern of rainfall. Other than that, according to [24-25] who had carried out three rainfall forecasting models based on a monthly data basis where he used ARIMA, ANN and MLR techniques to analyses his research. The results that he obtained was the application of neural network model were more capable of forecasting an accurate result compared to the other two models.

A study from [26] proposed a dynamic recurrent time-delay neural network for monthly rainfall forecast in Queensland, Australia. The network prototype models have a lower error when compared with the forecasts generated by the standard models used by the Australian Bureau of Meteorology. Moreover, they carried out three rainfall forecasting models that were developed based on ARIMA, ANN and multiple linear regression. The rainfall was estimated based on a monthly basis. They observe that the multilayer feed forward Back-Propagation (BP) neural network model forecast was better than the other two models.

The aim of this study is to compare which method can provide more accurate forecasting results for daily rainfall data. The data are analyzed by using two methods which are ARIMA and ANN methods. The models obtained from each method will be evaluated and compared based on the forecasting performances measurement of Mean Absolute Error, Mean Forecast Error, Root Mean Square Error and coefficient of determination. The models that provides a high accuracy will be chosen as the best model to forecast daily rainfall.

## 2. RESEARCH METHOD

There are several methods that can be used to forecast rainfall data. In this study, two methods were applied to forecast rainfall data which are ARIMA model and ANN model. Firstly, the data need to undergoes data pre-processing before proceeding with any time series method. These techniques are involved with data normalization, data lagging and data splitting.

Data normalization is one of methods used in data pre-processing to obtain the precision of the forecast of the model. The range of data used in this method were between [0,1]. The purpose to use this normalization range is because sigmoid (logistic) activation function was used in this study. Next, for data lagging is where it involves with layers of nodes which consist of input nodes, hidden layers and output nodes. In this research, the input variables are constructed through trial and error method.

The data are need to be splitting into training and testing set. In data splitting, training set will have more allocation than testing set with a proportion of 90% versus 10%, 80% versus 20% or 70%

## 2.1. Study area

Figure 1. Map of the study area

## 2.2. ARIMA model

The first step in model identification are to determine the time series is stationary or not stationary. If the data is not stationary, a non-seasonal differencing can be applied to the data to make it stationary. After that, the models can be identified according to the guideline of the autocorrelations functions (ACF) plot and partial autocorrelations function (PACF) plot. The guideline for the model identification are shown in Table 1.

Table 1. Guideline for ARIMA model identification

ACF	PACF	Model
Dies down	Cuts off after lag $p$	AR( $p$ )
Cuts off after lag $q$	Dies down	MA( $q$ )
Dies down	Dies down	ARMA( $p, q$ )
Cuts off after lag $q$	Cuts off after lag $p$	AR( $p$ ) or MA( $q$ )

Next, after the models have been identified, the estimation of the constant and coefficients of the equation must be obtained. For model estimation, we need to estimate the parameter for a tentative model

that has been selected. Parameters that are judged significantly different from zero are retained in the fitted model while parameters that are not significant are dropped from the model.

Last but not least, the adequacy of the model must be checked in model diagnostic checking step. Box-Ljung test is used for testing the lack of fit of a time series model and residual of the time series are correlated or uncorrelated after fitting an ARIMA model to the data. The model with the smallest  $p$ -value of the estimated parameter value and the highest  $p$ -value of the Box-Ljung test was chosen. Other than that, the model is also being selected using the Akaike Information Criterion (AIC) criteria. The model with the smallest AIC value will be chosen which shows an adequate model.

### 2.3. ANN model

ANN is one of the other methods that can be applied in time series analysis which is widely used by the researchers. There are several steps that are required in order to successfully forecast the neural networks model. The steps are network architecture, learning algorithm and the activation functions. Data normalization are often performed before executing with the ANN model in training process. The input data need to be normalized according to the activation function used. It can help to minimize the error of the model.

In this study, MLP structure consists of three layers which are input layer, hidden layer and output layer. A total of 35 MLP network models were developed using daily rainfall data. For input nodes, it is determined according to data lagging technique. The application of using data lagging technique was to evaluate the forecasting performance of the model in details and the capability of the models. The generated lagging observations were obtained from trial and error of input variables. The training number of hidden layer nodes used in this modelling are from 2 to 10 which is based on previous researcher's paper and also through trial and error. Furthermore, there is only one node used for output layer.

In modelling of ANN model for daily rainfall data, the network that applied was MLP which contains input, hidden and output layer. The models were trained based on learning method which is the gradient descent back-propagation algorithm. This algorithm consists of two parameters which are learning rate ( $lr$ ) and momentum coefficient ( $mc$ ). The parameters were determined based on previous researcher's paper and through trial and error method. The neural network model was trained with  $lr$  of 0.3, and  $mc$  parameters of 0.2 and number of training epochs was 1000.

For the activation function, two activation function were needed to link the neurons. For this study, Sigmoid (logistic) activation function was used for the hidden layer. This activation function keeps the range for the hidden layers to be within 0 to 1. Next, linear activation function was used at the output layer as there is only one result that is generated at the output layer, so the used of linear function is acceptable. Both of the equation are as follows:

$$f(x) = \text{purelin}(x) = x \quad (1)$$

$$f(x) = \log \text{sig}(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

where  $x$  is the input value.

### 2.4. Forecasting performance measurement

The performances of the models are evaluated by calculating difference between the observed rainfall data and the model generated rainfall data. According to [30-32], there are several performance evaluation methods which could be used for hydrological forecasting model. For this study, the forecasting performance is evaluated by using Mean Absolute Error (MAE), Mean Forecast Error (MFE), Root Mean Square Error (RMSE) and coefficient of determination ( $R^2$ ). The forecasting model that provides the smallest value of MAE, MFE and RMSE were appointed as the best model for forecasting. In addition, for the value for  $R^2$  which are between 0 and 1 were chosen which shows how well the data can fit the model. The formula is shown below:

$$\text{MAE} = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (4)$$

$$MFE = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)}{n} \quad (5)$$

$$R = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}} \quad (6)$$

$$R^2 = \left( \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}} \right)^2 \quad (7)$$

where  $y_t$  is the observed value at period  $t$ ;  $\hat{y}_t$  is predicted value at period  $t$ ;  $n$  is the number of periods used in calculation;  $R$  is the correlation coefficient;  $R^2$  is the coefficient of determination;  $n$  is the number of pairs of data;  $x$  is the observed value of rainfall data;  $y$  is the predicted value of rainfall data.

### 3. RESULTS AND DISCUSSION

The daily rainfall data sets were successfully forecasted using both ARIMA and ANN models. ARIMA (3,1,1) have the significant  $p$ -value for each of the parameters which is less than the significant level of  $\alpha = 0.05$ . Moreover, the smallest AIC value and largest Box-Ljung test are being selected with the value of 11060.89 and 0.2313 as the best model to forecast daily rainfall in Simpang Ampat, Pulau Pinang in ARIMA method.

Furthermore, ANN model is capable of predicting and efficiently as this method were involve in a nonlinear modelling of rainfall data. The ANN structure consists of seven input nodes, two to ten hidden layer nodes and one output nodes. A Feed Forward Back-Propagation Neural Network of ANN model was developed where the model was trained based on gradient descent back-propagation algorithm with sigmoid (logistic) activation function for hidden layer and linear activation function for output layer. ANN (6,4,1) model was trained and tested.

Next, the forecasting accuracy was measured according to observed and predicted value of the models. ARIMA and ANN models were evaluated and compared to see which method provides the most appropriate forecasting tools to forecast daily rainfall data. There is a difference based on the results obtained from accuracy checking between ARIMA and ANN models. The measurement of the level of accuracy is based on MAE, MFE, RMSE and  $R^2$  criteria.

From the analysis results, both of the models are capable as a forecasting tools to forecast daily rainfall data. The models that provides the smallest error of MAE, MFE, RMSE and the highest  $R^2$  were appointed as the best forecasting models. Table 2 shows the comparison of the performance measure for ARIMA and ANN models. Based on the results, the performance measure for ANN model shows a better result compared to ARIMA model where ANN model outperforms ARIMA model. The error measure of the training and testing set for MAE in ARIMA model is 78.0877 and 102.7644 respectively. The error measure of ANN model has a lower value of error measure in both training and testing set of MAE at 25.6573 and 8.4208.

While for MFE, if the value of MFE is larger than 0, it indicates the model is under-forecast. If it is less than 0, the model may lead to over-forecast. The MFE value for training and testing of ARIMA model were 2.9465 and -3.3436 which indicates that the model is under-forecast during training the model and tends to be over-forecast in testing set. This means that forecast for ARIMA model to be low in relation to the actual demand in training set. While for testing set, the forecast value is high in relation to actual demand. However, for ANN model, undergoes under-forecast in both training and testing set. The model tends to be under-forecast with an average absolute error of 13.1511 in training and 2.2188 in testing which the forecast has been in a low relation to the actual demand.

The two models were being compared based on RMSE value where it is used to compute the variations of the observed values and the predicted values. The smaller the value of RMSE, the better and accurate the results of the forecast would be. The RMSE value for ANN model are lower, while for ARIMA model is higher. This means that ANN model can provide more accurate results in forecasting daily rainfall data.

The coefficient of determination,  $R^2$  can also be measured and compared. This measurement error helps to see whether it gives the fluctuations percentages of one variables that is predicted from other variables. The value of  $R^2$  is between 0 and 1 where it denotes the strength of a linear association between x and y. In addition, it represents the percent of data that is closest to the line of best fit. The closer its value to 1, the better the fit or relationship between the variables. The results of ANN model for training and testing set shows the value of  $R^2$  is between 0 and 1 that is high at 0.8227 and 0.9432. ANN model shows a better fit and positive relationship between the variables compared to ARIMA model. Hence, ANN model outperforms ARIMA model and proven to model and forecast daily rainfall data. Figure 2 displayed a graph of comparison of ARIMA model versus ANN model.

Table 2. Comparison of the performance measure of selected ARIMA and ANN models

Method	Error Measure for Daily Rainfall Data in Training Set				Error Measure for Daily Rainfall Data in Testing Set			
	MAE	MFE	RMSE	$R^2$	MAE	MFE	RMSE	$R^2$
ARIMA (3,1,1)	78.0877	2.9465	131.1353	0.1231	102.7644	-3.3436	138.9895	0.0144
<b>ANN (6,4,1)</b>	<b>25.6573</b>	<b>13.1511</b>	<b>57.2931</b>	<b>0.8227</b>	<b>8.4208</b>	<b>2.2188</b>	<b>34.6740</b>	<b>0.9432</b>

Notes: Bold is the smallest value.

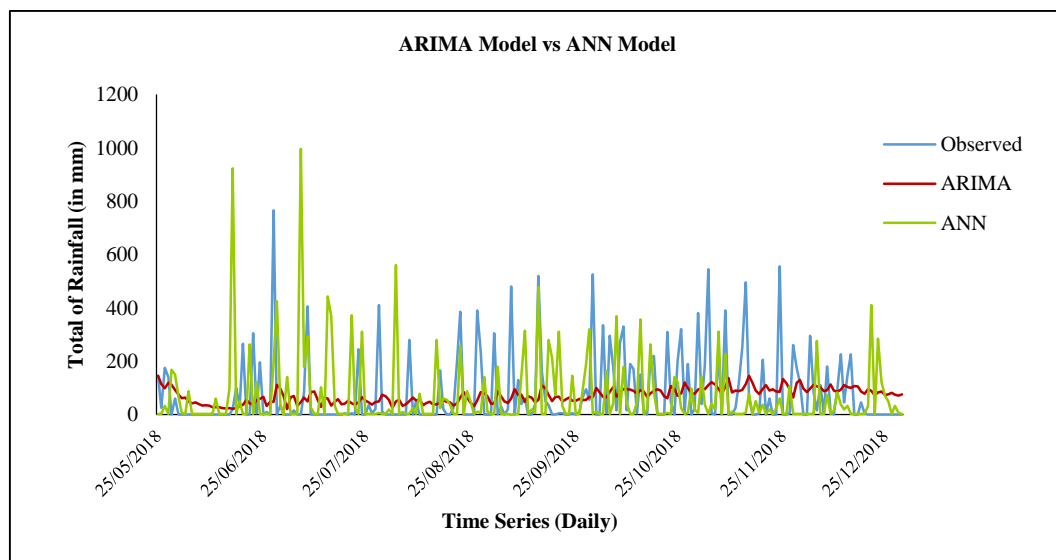


Figure 2. Comparison of ARIMA model vs ANN model

#### 4. CONCLUSION

Based on the results obtained, the two models are compared and evaluated in order to find the best forecasting model to forecast rainfall data. The purpose of comparing these two models was to find the most suitable method to forecast daily rainfall data which contains the minimum accuracy measure. For this comparison, there are three types of error measure that was being used to evaluate the accuracy measure of the models which are MAE, MFE and RMSE. The smaller the error, the more accurate the forecasted results of the models would achieve. Moreover, the coefficient of determination,  $R^2$  were also being measured as to inspect the better fit or relationship between the variables. The model that has the lowest error and  $R^2$  closest value to 1 for better fit of the variables are selected as the best model.

The ANN (6,4,1) shows the MFE value of 2.2188 that is smaller compared to ARIMA (3,1,1). It shows that ANN model is under-forecast which the forecasted results has been in a low relation to actual demand of 2.2188 mm in average for that day to rain. While, MFE for ARIMA (3,1,1) has the value of

-3.3436 mm in average which gives the meaning that the forecasted value for ARIMA model are too high to actual demand which tends to be over-forecast the results. Thus, ANN model was chosen as the best model that gives the lowest error on MAE, MFE and RMSE compared to ARIMA. In addition, a higher value of  $R^2$  which fitted the model to the variables much better.

## ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Education Malaysia for supporting this research under Fundamental Research Grant Scheme Vot No. FRGS/1/2018/STG06/UTHM/03/3 and partially sponsored by Universiti Tun Hussein Onn Malaysia.

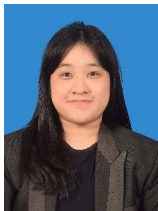
## REFERENCES

- [1] N. Q. Hung, M. S. Babel, S. Weesakul, & N. K. Tripathi, An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *Hydrology and Earth System Sciences*, 13(8), 1413-1425, 2009.
- [2] W. Wang, W. *Stochasticity, Nonlinearity and Forecasting of Streamflow Processes*. IOS press, Amsterdam, 2006.
- [3] D. Hirani & N. Mishra, A survey on rainfall prediction techniques. *International Journal of Computer Application*, 6(2), 28-42, 2006.
- [4] V. P. Singh & P. K. Chowdhury, Comparing some methods of estimating mean a real rainfall. *JAWRA Journal of the American Water Resources Association*, 22(2), 275-282, 1986.
- [5] M. M. Poulton, Neural networks as an intelligence amplification tool: A review of applications. *Geophysics*, 67(3), 979-993, 2002.
- [6] J. F. Kaashoek & H. Van Dijk, H, Long term values of euro/dollar and European exchange rates: A neural network analysis. *Medium Econometrische Toepassingen*, 10(4), 26-29, 2001.
- [7] F. M. Tseng, H. C. Yu & G. H. Tzeng, Combining neural network model with seasonal time series ARIMA model, *Technological Forecasting & Social Change*, 69, 71-87, 2002.
- [8] H. Altun, A. Bilgil & B. C. Fidan, Treatment of multi-dimensional data to enhance neural network estimators in regression problems. *Expert Systems with Applications*, 32(2), 599-605, 2007.
- [9] G. A. Fallah-Ghahary, Annual rainfall forecasting by using mamdani fuzzy inference system. *Research Journal of Environmental Sciences*, 3(4), 400-413, 2009.
- [10] C. L. Wu, K. W. Chau & C. Fan, Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *Journal of Hydrology*, 389(1), 146-167, 2010.
- [11] A. H. El-Shafie, A. El-Shafie, H. G. El Mazoghi, A. Shehata & M. R. Taha, Artificial neural network technique for rainfall forecasting applied to Alexandria, Egypt. *International Journal of Physical Sciences*, 6(6), 1306-1316, 2011.
- [12] O. KİŞİ, Daily river flow forecasting using artificial neural networks and auto-regressive models. *Turkish Journal of Engineering and Environmental Sciences*, 29(1), 9-20, 2005.
- [13] V. K. Somvanshi, O.P. Pandey, P.K. Agrawal, N. V. Kalanker, M. R. Prakash & R. Chand, Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. *J. Ind. Geophys. Union*, 10(2), pp. 141-151, 2006.
- [14] M. Castellano-Méndez, W. González-Manteiga, M. Febrero-Bande, J. M. Prada-Sánchez & R. Lozano-Calderón, Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box-Jenkins and neural networks methods. *Journal of Hydrology*, 296(1-4), 38-58, 2004.
- [15] Sutikno, H. A. Wibawa, P. Y. Budiart, "Classification of Road Damage from Digital Image Using Backpropagation Neural Network", *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 6, no. 4, 2017.
- [16] N. A. Sulaiman, M. P. Abdullah, H. Abdullah, M. N. S. Zainudin, A. M. Yusop, "Fault detection for air conditioning system using machine learning", *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 1, 109-116, 2020.
- [17] S. Verma, G.T. Thampi, M. Rao, "ANN based method for improving gold price forecasting accuracy through modified gradient descent methods", *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 1, 46 - 57, 2020.
- [18] Mohammed Sarhan Al Duais et al., "Improved Time Training with Accuracy of Batch Back Propagation Algorithm Via Dynamic Learning Rate and Dynamic Momentum Factor", *IAES International Journal of Artificial Intelligence (IJAI)*, vol. /issue 7(4), pp. 170-178, 2018.
- [19] T. Partal & H. K. Cigizoglu, Prediction of daily precipitation using wavelet-neural networks. *Hydrological sciences journal*, 54(2), 234-246, 2009.
- [20] R. V. Ramana, B. Krishna, S. R. Kumar & N. G. Pandey, Monthly rainfall prediction using wavelet neural network analysis. *Water resources management*, 27(10), 3697-3711, 2013.
- [21] K. K. Kuok, *Parameter Optimization Methods for Calibrating Tank Model and Neural Network Model for Rainfall-runoff Modeling* (Doctoral dissertation, Ph. D. Thesis. University Technology Malaysia), 2010.
- [22] M. A. Sharma, Comparative study of Rainfall forecasting models. *New York Science Journal*, 4(7), 115-120, 2011.
- [23] M. Khashei & M. Bijari, A new hybrid methodology for nonlinear time series forecasting. *Modelling and Simulation in Engineering*, 15, 2011.



- [24] S. A. Muttaleb, Prediction of Monthly Rainfall in Kirkuk Using Artificial Neural Network and Time Series Models. *Journal of Engineering and Sustainable Development*, 18(1), 129-143, 2014.
- [25] S. Barhmi & O. El Fatni, Hourly wind speed forecasting based on Support Vector Machine and Artificial Neural Networks. *IAES International Journal of Artificial Intelligence*, 8(3), 286, 2019.
- [26] J. Abbot & J. Marohasy, Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Advances in Atmospheric Sciences*, 29(4), 717-730, 2012.
- [27] S. Ismail, A. Shabri, R. Samsudin, "A hybrid model of self organizing maps and least square support vector machine for river flow forecasting", *Hydrology and Earth System Sciences*, 16, 4417-4433, 2012.
- [28] R. J. Hyndman & G. Athanasopoulos, "Forecasting: Principles and Practice," Otexts, 2012.
- [29] Box, G. E. P. G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 1976.
- [30] H. Karim, S. R. Niakan & R. Safdari, Comparison of neural network training algorithms for classification of heart diseases. *IAES International Journal of Artificial Intelligence*, 7(4), 185, 2018.
- [31] W. Wang, K. Chau, C. Cheng & L. Qiu, A Comparison of Performance of Several Artificial Intelligence Methods for Forecasting Monthly Discharge Time Series. *Journal of Hydrology*, 374(3-4), 294-306, 2009.
- [32] R. S. Govindaraju & A. R. Rao, Artificial neural Networks in hydrology part I. *Journal of Hydrologic Engineering*, 5(2), 115-123, 2002.

## BIOGRAPHIES OF AUTHORS



Mazwin Arleena Masngut was born in Penang, Malaysia, in 1997. She received the secondary education in Science Stream from Sekolah Menengah Kebangsaan Seri Nibong, Nibong Tebal, Penang. After that, she continued her studies for Physical Science at Matriculation which is in Penang Matriculation College, Kepala Batas, Penang, in 2016. She received her bachelor's degree in Industrial Statistics from UTHM, in 2020. She hopes that she can become an excellent data analyst in the future.



Shuhaida Ismail is a lecturer at the Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia (UTHM). She obtained her first degree in Computer Sciences majoring from UTM. She also obtained a Master degree and PhD from the same university. Throughout her studies, she developed an interest in Machine Learning research area, specifically in predictive modelling, classification, and clustering. Her current research areas are in Hydrological Modelling, Big Data Analytics and Deep Learning.



Aida Mustapha received the B.Sc. degree in Computer Science from Michigan Technological University and the M.IT degree in Computer Science from UKM, Malaysia in 1998 and 2004, respectively. She received her Ph.D. in Artificial Intelligence focusing on dialogue systems. She is currently an active researcher in the area of Computational Linguistics, Soft Computing, Data Mining, and Agent-based Systems.



Suhaila Mohd. Yasin obtained her first degree in Computer Science and her Master of Science (Computer) from UTM. She recently obtained her Ph.D. in Computer Science focusing on Software Testing from The University of Queensland. Her current research area is in search-based and model-based testing.